Towards the Fusion of Gaze and Micro-Gestures

Kagan Taylor*
Lancaster University

Haopeng Wang†
Lancaster University

Florian Weidner‡
Lancaster University

Hans Gellersen§
Lancaster University,
Aarhus University

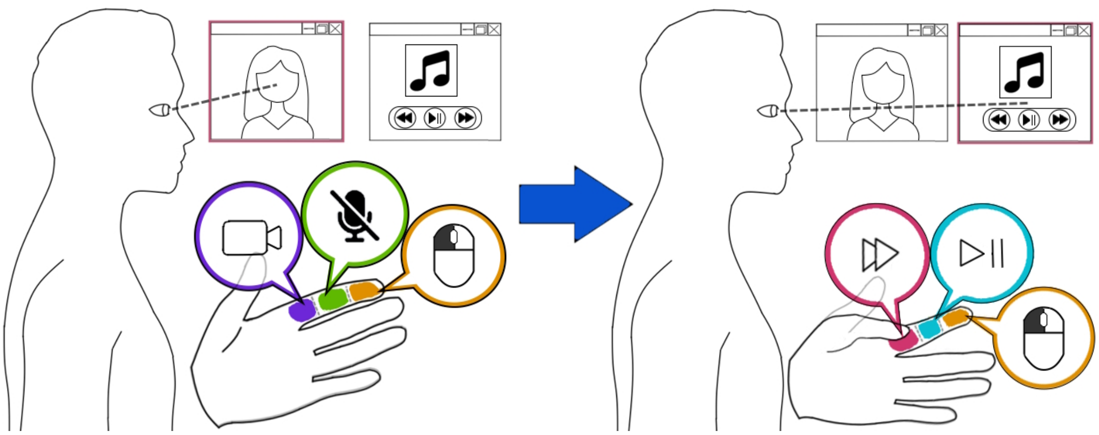

Figure 1: With gaze and micro-gestures, the gaze defines the target that receives input while thumb-to-finger-segment pinches provide the input. This figure shows an example in a virtual desktop setting. The fingertip is static, mimicking a mouse button: its functionality does not change, regardless of gaze location. The other two pinch locations change depending on where the user looks. The user controls their camera and microphone when looking at their video call. When looking at their music player, they control it — without having to fixate on the buttons.

ABSTRACT

Gaze and hand-based micro-gestures have been widely studied individually, each offering unique advantages as input modalities, but their combination remains under-explored. In this work, we integrate gaze with thumb-to-finger micro-gestures, discussing their complementary strengths to enable expressive and efficient interactions. We examine two strategies for combining gaze and micro-gestures, assess their strengths and weaknesses, and present examples such as mouse emulation and application shortcuts. Finally, we outline the potential and challenges of gaze and micro-gestures to enhance interaction across diverse contexts, and discuss future directions. This work begins the exploration of gaze and micro-gestures for novel input techniques.

Index Terms: Human-computer interaction, micro-gestures, gaze, virtual reality, augmented reality, head-mounted displays.

1 INTRODUCTION

With the introduction of the Apple Vision Pro, Gaze+Pinch [4] became a prominent interaction technique, leveraging both the pointing speed of gaze and the efficient and available input possible with the pinch (a thumb-to-index micro-gesture). By introducing additional hand-based micro-gestures, we can extend the applicability of this interaction paradigm.

Performing a pinch on finger segments other than the fingertip is particularly well-suited for interaction, as the finger segments provide clear proprioceptive landmarks [7]. This allows users to perform gestures without looking at the hand [2], leaving the gaze free to explicitly select content or implicitly provide context. Despite the popularity of Gaze+Pinch, and the great attention that both gaze [5] and micro-gestures [1, 6] have received, existing works have yet to fully explore their combined potential.

In this paper, we explore the relationship between gaze and micro-gestures, examining two strategies for combining them, and discuss each approach's advantages and disadvantages. We further present two applications of gaze and micro-gestures, highlighting their potential through examples such as mouse emulation and keyboard shortcuts. Finally, we discuss the future directions for this interaction paradigm. With this work, we seek to inspire discussion about the combination of gaze and micro-gestures and motivate further research.

2 COMBINING GAZE AND MICRO-GESTURES

We identify two complementary strategies to combine gaze and micro-gestures. One way to implement micro-gestures is to define them as static gestures that deliver functionality to the precise gaze location. This approach mimics the functionality of a mouse, where predefined actions are applied at the cursor's position. Such an approach is particularly effective for issuing generic commands across multiple objects and widgets, and is employed in existing gaze + micro-gesture works such as Gaze+Pinch [4, 3]. Examples of this approach are, for example, the left- and right-click of a mouse or other input techniques that have a fixed system-wide meaning, such as 'Ctrl+Z' for undo.

Alternatively, micro-gestures can be assigned dynamically, leveraging the wider context of gaze. This strategy is analogous to established interaction paradigms, such as controller buttons adapting their functionality depending on the video game being played or function keys of a keyboard altering their roles based on the application in focus. In this approach, the gaze defines the focus (e.g.,

*e-mail: k.taylor11@lancaster.ac.uk
†e-mail: h.wang73@lancaster.ac.uk
‡e-mail: f.weidner@lancaster.ac.uk
§e-mail: h.gellersen@lancaster.ac.uk

which application receives input), allowing almost instantaneous context switching. For instance, when focusing on a music player, a micro-gesture's function could dynamically change from the default 'select' to 'play/pause'. This offers a more efficient and adaptable system; however, context-defined micro-gestures can be more difficult to learn and discover, especially across many contexts.

These two strategies can also be combined: essential universal micro-gestures can be statically defined to preserve their broad applicability, while additional micro-gestures can be dynamically assigned for context-specific functionality. For instance, the tip of the index finger can be statically defined to always act as 'left click,' while gestures to the other finger segments are defined dynamically to provide context-specific functionality for more task-relevant interactions, as illustrated in Figure 1.

3 Applications of Gaze and Micro-Gestures

In the following, we discuss use cases that rely on gaze and micro-gestures to outline the working principle.

3.1 Mouse Emulation in VR

In Gaze+Pinch, the gaze controls the cursor position, while a static thumb-to-index finger micro-gesture confirms selections, mirroring the left-click function of a mouse. We propose an extension of this concept by adding two additional statically defined thumb-to-finger micro-gestures: one located on the middle phalanx and the other on the proximal phalanx of the index finger. This enables the functionality of the middle and right mouse buttons. Here, the finger segments intuitively map to the mouse buttons.

The system benefits from the improved pointing speeds offered by gaze, and using micro-gestures ensures the hands remain free, eliminating the need for additional devices (alongside the head-mounted display). However, using gaze to control the mouse position couples the user's visual attention with their input control. This alignment can enhance interaction efficiency when users focus on a target they intend to interact with, but unintended interactions can happen when they only intend to browse content.

3.2 Application Shortcuts

While simple, a single, statically defined micro-gesture (such as in Gaze+Pinch) affords only one means of interaction, limiting functionality to pointer-based tasks. This restriction necessitates frequent gaze movements to access various controls via buttons and menus. To combat this, a system can use dynamically defined thumb-to-finger micro-gestures using the middle and proximal phalanges of the index finger. These gestures can then serve as shortcuts or macros for applications, reducing the need for excessive pointing and thereby streamlining interaction. Figure 1 showcases an example where the middle and proximal phalanges take over application-specific functionality such as play/pause and skip forward for a music player but also mute/unmute and camera on/off for a video conference, depending on which window is the user looks at.

Dynamically defined micro-gestures would enable users to interact with applications by gazing at any part of the window rather than requiring precise selections of specific buttons and widgets. Being dynamic, novice users may struggle to discover a micro-gesture's functionality, leading to errors and unintended interactions. Additionally, the system may not scale well to include all possible thumb-to-finger gestures, as memorability would decline.

4 Challenges of Gaze and Micro-Gestures

Combining gaze and micro-gestures presents several challenges. While gaze tracking is reasonably accurate in commercial devices, tracking hand and finger movements remains difficult. Thus, advanced tracking technologies and algorithms are needed for robust micro-gesture recognition. In addition, the dynamic and static assignment of micro-gestures enables diverse use cases but can cause mode confusion, where users forget a gesture's function. Effective feedback that does not disrupt the interaction should be displayed in the world or on the hand to mitigate this. Also, overloading the hand with too many micro-gestures can impair memorability and learnability. To address this, assignments should align with user profiles (novice vs. expert) and focus on frequently used tasks [2]. Lastly, enabling always-on interaction can result in accidental activations. Improved tracking algorithms and context-aware systems (e.g., task understanding, location) are necessary to minimise such issues.

5 Conclusion

This paper explored the combination of gaze and micro-gestures as a novel interaction paradigm. We identified two strategies to combine gaze and multiple micro-gestures: directing micro-gesture inputs to the gaze point (static), or using gaze-defined context to set micro-gesture functions (dynamic). We outlined application scenarios and discussed research challenges.

Combining gaze and micro-gestures offers a promising foundation for future interaction techniques across various contexts, alleviating the need for physical interaction devices and making HMDs ready for interaction on the go. To do this, research must explore possible combinations of gaze and micro-gestures in user studies, focusing on qualitative and quantitative measures and overcome technological as well as user interface design challenges.

Acknowledgments

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant No. 101021229, GEMINI: Gaze and Eye Movement in Interaction).

References

[1] K. Danyluk, S. Klueber, A. S. Nittala, and W. Willett. Understanding gesture and microgesture inputs for augmented reality maps. In *Proceedings of the 2024 ACM ACM DIS Conference*, DIS '24, p. 409–423. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3643834.3661630 1

[2] S. A. M. Faleel, S. Kwon, D. Ahlstrom, and P. Irani. Validating Eyes-free Affordance of On-Finger Hand Proximate User Interfaces in In-situ Scenarios . In *2024 IEEE ISMAR*, pp. 1283–1292. IEEE Computer Society, Los Alamitos, CA, USA, Oct. 2024. doi: 10.1109/ISMAR62088.2024.00145 1, 2

[3] S. Pei, D. Kim, A. Olwal, Y. Zhang, and R. Du. Ui mobility control in xr: Switching ui positionings between static, dynamic, and self entities. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3613904.3642220 1

[4] K. Pfeuffer, B. Mayer, D. Mardanbegi, and H. Gellersen. Gaze + pinch interaction in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, SUI '17, p. 99–108. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3131277.3132180 1

[5] A. Plopski, T. Hirzle, N. Norouzi, L. Qian, G. Bruder, and T. Langlotz. The eye in extended reality: A survey on gaze interaction and eye tracking in head-worn extended reality. *ACM Comput. Surv.*, 55(3), Mar. 2022. doi: 10.1145/3491207 1

[6] A. Sharma, A. Ivanov, F. Lai, T. Grossman, and S. Santosa. Graspui: Seamlessly integrating object-centric gestures within the seven phases of grasping. In *Proceedings of the 2024 ACM DIS Conference*, DIS '24, p. 1275–1289. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3643834.3661551 1

[7] J. Steimle, J. Bergstrom-Lehtovirta, M. Weigel, A. S. Nittala, S. Boring, A. Olwal, and K. Hornbæk. On-skin interaction using body landmarks. *Computer*, 50(10):19–27, 2017. doi: 10.1109/MC.2017.3641636 1